

Data Discretization Process Of Health Management System

Tang Xiaokang,Zhang Xuezhi,Zou Qiong,Xie Cun

Wuhan Mechanical Technology Colledge, Wuhan ,430075, China

Keywords: Health Management System ; Discretization; Conditional entropy.

Abstract. The algorithm of attribute reduction based on discernibility matrix is advanced. The advanced algorithm uses the dependence of decision attribute to condition attribute as heuristic information. The advanced algorithm can reduce the complexity of discernibility matrix constructing effectively. The value reduction algorithm based on discernibility matrix is advanced, and wrong rules can be avoided.

The basic concept of rough set

As a new mathematical tool, Rough set theory can effectively deal with the fuzzy and imprecise problems, mining and processing the data in the complex system information. Its main idea is to remove redundant information system, finally complete the extraction of rules, realize the classification of the system provided by reduction without any prior knowledge based on the existing data. But rough set cannot be directly used for dealing with continuous attributes, this largely limits its application scope, continuous attribute discretization become an important aspect of application research on rough set theory, based on the concept of information entropy in this paper, a continuous attribute discretization method be put forward based on condition entropy, and validate the correctness of the method through an example.

Defined 1[1] given approximation space $K = (U, S)$, Among them, U is the theory of domain, S denote the equivalence relation of cluster in domain, $\forall X \subseteq U$ and U formate a equivalence relation $R \in IND(K)$, denote subset X about lower approximation and upper approximation of knowledge R respectively

$$\underline{R}(X) = \{x | (\forall x \in U) \wedge ([x]_R \subseteq X)\} = \bigcup \{Y | (\forall Y \in U / R) \wedge (Y \subseteq X)\} \quad (1)$$

$$\bar{R}(X) = \{x | (\forall x \in U) \wedge ([x]_R \cap X \neq \emptyset)\} = \bigcup \{Y | (\forall Y \in U / R) \wedge (Y \cap X \neq \emptyset)\} \quad (2)$$

Collection $bn_R(X) = \bar{R}(X) - \underline{R}(X)$ called R edge boundaries of X , it is the collection of elements which could not be judged belong to or not belong to the elements of U in the domain X by according to the knowledge R ; $pos_R(X) = \underline{R}(X)$ called R positive region of X , the collection consist of lower approximation $\underline{R}(X)$ and positive region $pos_R(X)$ whether must belong to or maybe belong to the element of U in X domain of discourse according to the knowledge R , the collection consist of upper approximation $pos_R(X)$ whether must belong to or maybe belong to the element of U in X domain of discourse according to the knowledge R , obviously $\bar{R}(X) = pos_R(X) \cup bn_R(X)$; $neg_R(X) = U - \bar{R}(X)$ called R negative domain of X , it is the collection consist of element of theory domain which not belong to X according to knowledge R . Collection of upper approximation, lower approximation and edge boundary can be represented.

Definition 2 given to theory domain and an equivalence relation R , $\forall X \subseteq U$, if $\underline{R}(X) = \bar{R}(X)$, collection X is precise set of R - or defined set of R - about theory domain U relative to the knowledge R ; if $\underline{R}(X) \neq \bar{R}(X)$, collection X is rough set of R - or undefined set of R - about theory domain U relative to the knowledge R . In the approximate space $K = (U, S)$, when

$\forall R \in IND(K)$, X is the rough set of R —, called rough set or undefined set in the approximate space. for the rough set, its lower approximate $\underline{R}(X)$ described the minimum defined set contained X , its upper approximate $\overline{R}(X)$ described the maximum defined set contained X .

Description if discretization problem

$S = (U, A, V, f)$ as a decision table, U is theory domain, $A = C \cup D$ is attribute set, $C = \{a \mid a \in C\}$ is condition attribute set, $D = \{d \mid d \in D\}$ is condition attribute set, and $C \neq \emptyset, D \neq \emptyset, C \cap D = \emptyset$, $V = \bigcup_{a \in A} V_a$ is attribute domain, f is information function, the breakpoint in the range values V_a of a attributes of a breakpoint can be referred to (a, c) , and $a \in A$, c is real number, the classification P_a is defined by an arbitrary breakpoints set $\{(a, c_1^a), (a, c_2^a), \dots, (a, c_k^a)\}$ in the range values $V_a = [l_a, r_a]$.

$$P_a = \{[c_0^a, c_1^a), [c_1^a, c_2^a), \dots, [c_k^a, c_{k+1}^a]\} \quad (3)$$

$$l_a = c_0^a < c_1^a < c_2^a < \dots < c_k^a < c_{k+1}^a = r_a \quad (4)$$

$$V_a = [c_0^a, c_1^a) \cup [c_1^a, c_2^a) \cup \dots \cup [c_k^a, c_{k+1}^a] \quad (5)$$

therefore, any $P = \bigcup_{a \in A} P_a$ defined a new decision table $S^p = (U, A, V^p, f^p)$, $f^p(x_a) = i \Leftrightarrow f(x_a) \in [c_i^a, c_{i+1}^a)$, for $x \in U$, $i \in \{0, \dots, k\}$, k is the cardinal number of P_a , the original decision table is replaced by a new decision table after discretization.

Continuous attribute discretization is essentially Insert some breakpoints in the scope of the range value of continuous attributes, continuous attributes are divided into several discrete intervals. If divided thinnly, can improve the attribute dependency, make decision table classification ability enhancement, but tend to increase the complexity, it is not good for attribute reduction; if ived coarsely, may lead to the increase of incompatible information in decision table1.

U as a theory domain, P and Q is the two equivalent relations in the U . X and Y separately as a exported division from the P and Q in the U , and

$$\begin{aligned} X &= U / IND(P) = \{X_1, X_2, \dots, X_n\} \\ Y &= U / IND(Q) = \{Y_1, Y_2, \dots, Y_m\} \end{aligned} \quad (6)$$

then:

(1) subset probability distribution of P and Q in the U separately is

$$[X; P] = \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ p(X_1) & p(X_2) & \dots & p(X_n) \end{bmatrix} \quad (7)$$

$$[Y; P] = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_m \\ p(Y_1) & p(Y_2) & \dots & p(Y_m) \end{bmatrix} \quad (8)$$

$$\text{and, } p(X_i) = \frac{\text{card}(X_i)}{\text{card}(U)}, (i = 1, 2, \dots, n), \quad p(Y_j) = \frac{\text{card}(Y_j)}{\text{card}(U)}, (j = 1, 2, \dots, m)$$

(2) Definition of joint probability distribution of P and Q

$$[XY; P] = \begin{bmatrix} X_1 \cap Y_1 & X_2 \cap Y_2 & \dots & X_n \cap Y_m \\ p(X_1 Y_1) & p(X_2 Y_2) & \dots & p(X_n Y_m) \end{bmatrix} \quad (9)$$

and, the formula of the probability density of product event is

$$p(X_i Y_j) = \frac{\text{card}(X_i \cap Y_j)}{\text{card}(U)}, (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (10)$$

Definition of information entropy and conditional entropy can get from probability distribution as follows:

Definition 2 given the knowledge P and its probability distribution, then

$$H(P) = -\sum_{i=1}^n p(X_i) \log_2 p(X_i) \quad (11)$$

is information entropy of knowledge P .

Definition 3 Given the knowledge P and Q and their respective probability and conditional probability distribution, then

$$H(Q|P) = -\sum_{i=1}^n p(X_i) \sum_{j=1}^m p(Y_j | X_i) \log_2 p(Y_j | X_i) \quad (12)$$

is the conditional entropy of knowledge Q relative to P .

$$\begin{aligned} S_2([0, 0.3]) \wedge S_4([0.1, 0.55]) \wedge S_8([0, 0.45]) \wedge S_9([0.675, 0.9]) \wedge S_{11}([0, 0.5]) &\rightarrow D(1) \\ S_2([0.6, 0.9]) \wedge S_4([0.55, 1]) \wedge S_8([0.45, 0.9]) \wedge S_9([0.45, 0.675]) \wedge S_{11}([0.5, 1]) &\rightarrow D(0) \\ S_2([0.6, 0.9]) \wedge S_4([0.55, 1]) \wedge S_8([0.45, 0.9]) \wedge S_9([0, 0.225]) \wedge S_{11}([0.5, 1]) &\rightarrow D(0) \\ S_2([0.3, 0.6]) \wedge S_4([0.55, 1]) \wedge S_8([0, 0.45]) \wedge S_9([0, 0.225]) \wedge S_{11}([0, 0.5]) &\rightarrow D(0) \\ S_2([0.6, 0.9]) \wedge S_4([0.55, 1]) \wedge S_8([0, 0.45]) \wedge S_9([0, 0.225]) \wedge S_{11}([0.5, 1]) &\rightarrow D(0) \\ S_2([0.3, 0.6]) \wedge S_4([0.55, 1]) \wedge S_8([0, 0.45]) \wedge S_9([0, 0.225]) \wedge S_{11}([0.5, 1]) &\rightarrow D(0) \\ S_2([0.3, 0.6]) \wedge S_4([0.1, 0.55]) \wedge S_8([0, 0.45]) \wedge S_9([0, 0.225]) \wedge S_{11}([0.5, 1]) &\rightarrow D(0) \\ S_2([0.6, 0.9]) \wedge S_4([0.55, 1]) \wedge S_8([0.45, 0.9]) \wedge S_9([0.225, 0.45]) \wedge S_{11}([0.5, 1]) &\rightarrow D(1) \end{aligned}$$

Algorithm implementation

Table1 the discretization results

Condition attribute	K	Discretization interval			
		1	2	3	4
S1	0	[0.2,0.8]			
S2	2	[0,0.3)	[0.3,0.6)	[0.6,0.9]	
S3	0	[0,0.9]			
S4	1	[0.1,0.55)	[0.55,1]		
S5	3	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
S6	0	[0,1]			
S7	3	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
S8	1	[0,0.45)	[0.45,0.9]		
S9	3	[0,0.225)	[0.225,0.45)	[0.45,0.675)	[0.675,0.9]
S10	3	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
S11	1	[0,0.5)	[0.5,1]		

Decision table is discretized and reduced according to the given algorithm. In this paper, test the test data by using the extracted rules, the test results as shown in table 2. It can be seen from the result of the experiment, the test data can be classified correctly, signify that the method has certain validity.

Table 2 test result

Test sample	S2	S4	S8	S9	S11	D
x1	[0,0.3)	[0.1,0.55)	[0,0.45]	[0.675,0.9]	[0,0.5)	1
x7	[0.6,0.9]	[0.55,1]	[0.45,0.9]	[0.45,0.675)	[0.5,1]	0
x11	[0.6,0.9]	[0.55,1]	[0.45,0.9]	[0,0.225)	[0.5,1]	0
x16	[0.6,0.9]	[0.55,1]	[0,0.45)	[0,0.225)	[0.5,1]	0
x18	[0.3,0.6)	[0.1,0.55)	[0,0.45)	[0,0.225)	[0.5,1]	0
x21	[0.6,0.9]	[0.55,1]	[0.45,0.9]	[0.225,0.45)	[0.5,1]	1

Adopt example in reference [4] to experiment as shown in table 3, the example has 11 condition attributes and 1 decision attribute, 21 groups data altogether.

Table 3 failure data

U	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀	S ₁₁	D
x	0.8	0	0.	0.1	1	0	1	0.	0.9	1	0	1
x	0.8	0	0.	0.1	0.	0	0.	0.	0.8	0.9	0	1
x	0.5	0	0.	0.1	0.	0	1	0.	0.7	0.9	0	1
x	0.8	0	0.	0.2	1	0	1	0.	0.9	1	0	1
x	0.5	0	0.	0.1	0.	0	0.	0.	0.7	0.9	0	0
x	0.5	0.	0	0.8	0	0.	0.	0.	0.5	0.1	0.8	0
x	0.5	0.	0	0.8	0	0.	0.	0.	0.5	0.2	0.9	0
x	0.6	0.	0	0.9	0	0.	0.	0.	0.4	0.1	0.7	0
x	0.4	0.	0	0.7	0	0.	0.	0.	0.3	0.1	0.6	1
x	0.3	0.	0	0.9	0	1	0	0.	0.1	0	0.9	0
x	0.2	0.	0	0.8	0	1	0	0.	0	0	0.8	0
x	0.2	0.	0	0.6	0	0.	0	0.	0	0	0.6	0
x	0.4	0.	0.	0.6	0	0	0.	0.	0.2	0	0.1	0
x	0.4	0.	0.	0.7	0.	0	0	0.	0.1	0	0.1	0
x	0.4	0.	0.	0.9	0	0.	0	0.	0.1	0	0.9	0
x	0.3	0.	0.	1	0	1	0	0.	0	0	1	0
x	0.3	0.	0.	1	0	1	0	0.	0	0	1	0
x	0.6	0.	0.	0.3	0.	0	0	0	0	0	0.6	0
x	0.7	0.	0.	0.3	0.	0	0	0	0	0	0.8	0
x	0.7	0.	0.	0.6	0	0	0.	0.	0.3	0	0.9	1
x	0.7	0.	0.	0.7	0	0	0.	0.	0.4	0	0.8	1

extract 15 groups (x_2 、 x_3 、 x_4 、 x_5 、 x_6 、 x_8 、 x_9 、 x_{10} 、 x_{12} 、 x_{13} 、 x_{14} 、 x_{15} 、 x_{17} 、 x_{19} 、 x_{20}) from the failure data in table 1 as the training sample, other 6 groups as test data, discrete the continuous attribute by the algorithm In this paper, obtain $\beta = 0.01$, $Ka = 3$, the breakpoint number K of every condition attribute which finally determined by the algorithm In this paper and the discretization results are shown in table 4, obtain the minimum reduction $\{S_2, S_4, S_8, S_9, S_{11}\}$ after reduce by the method adopt in reference:

Table 4 the discretization results

Condition attribute	K	Discretization interval			
		1	2	3	4
S ₁	0	[0.2,0.8]			
S ₂	2	[0,0.3)	[0.3,0.6)	[0.6,0.9]	
S ₃	0	[0,0.9]			
S ₄	1	[0.1,0.55)	[0.55,1]		
S ₅	3	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
S ₆	0	[0,1]			
S ₇	3	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
S ₈	1	[0,0.45)	[0.45,0.9]		
S ₉	3	[0,0.225)	[0.225,0.45)	[0.45,0.675)	[0.675,0.9]
S ₁₀	3	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
S ₁₁	1	[0,0.5)	[0.5,1]		

Feedback information with decision attribute of entropy condition attribute which based on regular conditional entropy of continuous attribute discretization algorithm, can more accurately reflect the dependency relation when the decision attribute center condition attribute. The advantages of measure discretization of the feedback information. This algorithm measure the advantage of discretization feedback information according to the regular condition entropy when the decision attribute center condition attribute, It can be better reflect in incompatible information system. But the algorithm adopt the interval division method to select the candidate breakpoint, need to set the initial divided points artificially, then the discretization quality is not high in some instance.

References

- [1] Miao tuo qian, Li guo dao. Rough set theory、Algorithm and application [M]. Beijing: Tsinghua university press, 2008
- [2] Wang J, Miao D Q. Analysis on attribute reduction strategies of rough set[J]. Journal of Computer Science and Technology, 1998, 13(2): 189-192.
- [3] Walczak B, Massart D L. Rough sets theory[J]. Chemometrics and Intelligent Laboratory System, 1999, 47(1): 1-16.
- [4] Hu Xiaohua, Cercone N. Learning in relational databases: a rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-338.
- [5] Aminian F, Aminian M. Fault Disgnosis of Analog Circuits Using Bayesian Neural Networks with Wavelet Transform as Preprocessor[J]. Journal of Electronic Testing, 2001, 17(1): 29-36.
- [6] Cristianini N, Shawe-taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods[M]. Cambridge University Press, 2000.
- [7] Boser B, Guyon I, Vapnik V. A Training Algorithm for Optimal Margin Classifiers[C]. Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA: 1992, 144-152.
- [8] Hu Xiaohua, Cercone N. Learning in relational databases: a rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-338.
- [9] Miao Duoqian, Wang Jue. An information-based algorithm for reduction of knowledge[C]. IEEE ICIPS'97, 1997, 1155-1158.
- [10] Liu H, Setiono R. Chi2: Feature selection and discretization of numeric attributes[C]. Proceedings of the 7th International Conference on Tools with Artificial Intelligence, Herndon, VA, USA, 1995, 388-391.

- [11]Dougherty J, Kohavi R, Sahami M. Supervised and unsupervised discretization of continuous features[C]. Proceedings of 12th ICML. Morgan Kaufmann publishers, 1995, 194-202.
- [12]Nguyen H. S.,Showron A. Quantization of Real Value Attributes-Rough Set and Boolean Reasoning Approach[C]. Rroc of the Second Joint Conference on Information Sciences, 1995, 34-37.
- [13]Wang J, Miao D Q. Analysis on attribute reduction strategies of rough set[J]. Journal of Computer Science and Technology, 1998, 13(2): 189-192.